

Discovery of Re-usable Architecture/Engineering Construction Knowledge Through Exploration of a Corporate Memory

Renate Fruchter,¹ Kushagra Saxena,² Peter Demian³

¹ ASCE Member, Director of Project Based Learning Laboratory, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305-4020, fruchter@stanford.edu

² M.S., Department of Management Science & Engineering, Stanford University, Stanford, CA 94305-4020, kushagra@stanford.edu

³ Ph.D Candidate, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305-4020, pdemian@stanford.edu

Summary

Re-using knowledge in architecture, engineering and construction (AEC) firms can lead to greater competitive advantage, improved designs, and more effective management of constructed facilities. This paper discusses the importance of exploration and discovery of reusable knowledge from a corporate archive as opposed to simple search and retrieval. We describe and illustrate through a scenario of use an exploration framework and prototype, CoMemTM that formalizes the added value of exploration in the process of knowledge reuse. We discuss two exploration activities: (i) *Breadth-Only overview exploration* that assist a user to rapidly localize pockets of re-usable knowledge from the large corporate archive and (ii) *Iterative breadth-depth exploration* that enables a user to identify those re-usable components of the Corporate Memory that may yield design issues that were not originally considered.

1. Introduction

Managing and reusing knowledge in architecture, engineering and construction firms can lead to greater competitive advantage, improved products, and more effective management of constructed facilities. However reuse often fails, since knowledge is not captured, it is captured out of context rendering it not reusable, or there are no formal mechanisms from both the information technology and organizational viewpoints to find, explore, and retrieve reusable knowledge. We define design knowledge reuse as the reuse of previously designed and built facilities, subsystems, or components, as well as the knowledge and expertise ingrained in these previous projects.

This paper builds on the notion of *knowledge in context* (Fruchter and Demian 2002). Knowledge in context is design knowledge as it occurs in a designer's personal memory: rich, detailed, and contextual. We argue that in order for knowledge to be reusable, the user should be able to see and explore the context in which this knowledge was originally created and interact with this rich content. We call a repository of such knowledge in context the corporate memory. We view knowledge reuse as a step in the knowledge life cycle. Knowledge is created as AEC practitioners who collaborate on projects. It is captured, indexed, and stored in an archive. At a later time, it is retrieved from the archive, explored, understood, and reused. Finally, as knowledge is reused it is refined and becomes more valuable. In this sense, the archive system acts as a knowledge refinery.

The digital age holds great promise to assist in content archival and retrieval. Nevertheless, most digital content management applications today offer few solutions to assist designers to find, explore, and understand content in context. State-of-practice document management solutions are limited to digital archives of formal documents logged by discipline, project phase, date, and originator. A typical corporate archive grows as the firm works on more projects. Consequently, it

becomes increasingly challenging for users to search and: (1) have a “birds’ eye view” of all the corporate projects with flags to relevant past project information pertinent to specific problems they work on, (2) explore specific items in context, and (3) discover reusable items or topics that were originally not considered.

Based on extensive ethnographic studies of AEC professionals at work we formalized two key activities in the process of knowledge reuse: *finding* reusable items and *understanding* these items in context, i.e., exploration of the context and the history of potentially reusable items (Fruchter and Demian 2002). We briefly describe the prototype, CoMem that supports these activities. In this paper, we emphasize the importance of exploration and discovery of reusable knowledge as opposed to simple search and retrieval. Specifically, we describe an exploration framework that formalizes the added value provided by CoMem™ for exploration and the knowledge reuse process as opposed to simple recall by experts or a random search. The framework further categorizes exploration activity in the following: (i) *Breadth-Only overview exploration* that helps localize pockets of re-usable knowledge from the large corporate archive and (ii) *Iterative breadth-depth exploration* that helps discover those re-usable components of the Corporate Memory that may not be contextually relevant to the original problem query. We believe, exploration activity combining both types, i.e., breadth-only and breadth-depth, will bring to the foreground design issues that the users may not have considered in their original query for re-usable knowledge and will result in a more detailed and informed solution. To demonstrate the generality of CoMem we will illustrate its integration with two different hierarchical archives, one consisting of a testbed with 10 building information model projects, and the second consisting of an archive of project discussion forums, called ThinkTank consisting of over 60 projects.

2. CoMem™ System Overview

CoMem™ (Fruchter and Demian 2002) is a tool developed at the PBL Lab to support the process of reuse of knowledge stored in corporate project design databases. CoMem is distinguished from the document-centric state-of-practice solutions by its approach of “overview first, zoom and filter, and then details-on-demand.” (Schneiderman 1994, Schneiderman 1999).

Based on the three reuse steps identified during the ethnographic study – *find, explore project context, explore evolution history* – CoMem has three corresponding modules: an *overview*, a *project context explorer*, and an *evolution history explorer*. (Figure 1) For each module, we investigated various radically different *metaphors*. We use *metaphor* here in a human-computer interaction sense. Metaphors increase the usability of user interfaces by supporting understanding by analogy. Modern operating systems use the desktop metaphor. Online services use shopping cart and checkout metaphors to relate the novel experience of buying online to the familiar experience of buying at a bricks and mortar store.

CoMem uses a *map metaphor* for the overview. The CoMem Overview Module provides a succinct “at a glance” view of the entire corporate memory. Corporate memories are usually hierarchical, where the corporation contains multiple projects, each project consists of multiple disciplines or building subsystems (structural, electrical, and so on), and each discipline contributes multiple components. This archive can become very large. For example, consider a small design firm that has worked on 10 projects, each project involving 10 disciplines, each discipline contributing 20 components, with each object versioned 50 times over the course of the project. The total number of objects in the corporate memory is close to 10^5 . Conventional techniques for visualizing hierarchies using nodes connected by links are inappropriate, given that the overview needs to show the entire corporate memory in a single display. CoMem uses the “*squarified treemap*” algorithm

to display an overview of the corporate memory in which the hierarchy is visualized as a series of nested rectangles (Schneiderman1999). The area on the map allocated to each item is based on a measure of how much knowledge this item encapsulates, i.e. how richly annotated it is, how many times it is versioned, how much external data is linked to it. Each item on the map is color-coded by a measure of relevance to the designer's current task. (Figure 1a) Currently, this relevance measure is based on textual analysis of the corporate memory using Latent Semantic Indexing algorithm (Landauer and Dumais 1995).

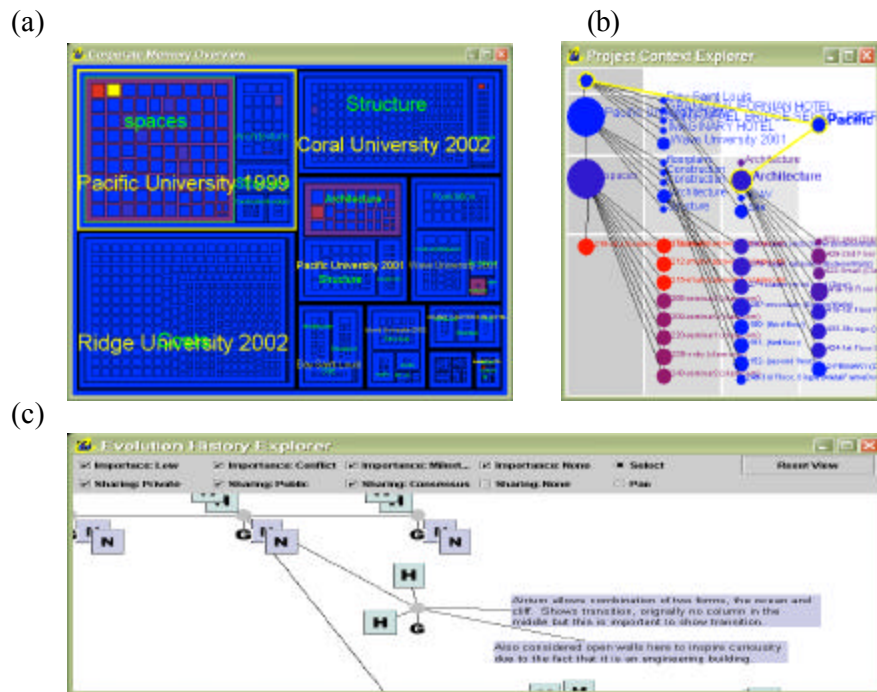


Figure 1: CoMem™ Modules: (a) The Map Overview provides a succinct snapshot of the entire corporate memory, in this case of engineering notes and models from a set of projects; (b) The Evolution History Explorer allows the user to explore the evolution of an item before reusing it; (c) The Project Context Explorer allows the user to explore related items in the corporate memory to better understand the item being reused

The CoMem Project Context Explorer (PCE) module enables the designer to explore the project context of any item selected from the overview. This item becomes the *focal point* of the interaction. CoMem uses a *fish-eye lens metaphor* for the project context explorer (Furnas 1981) (Lamping and Rao 1995). A fish-eye lens balances local detail with global context. This metaphor is used here to suggest that the designer is initially concerned only with the item of interest, but begins to explore the context “outwards” as necessary. Given a user-specified focal point, CoMem uses the *fish-eye formulation* to assign a degree of interest to every item in the corporate memory (Figure 1b)

Our observations show that the most striking means of transmitting knowledge from experts to novices in A/E/C design offices is through the informal recounting of experiences from past projects. Stories convey great amounts of knowledge and information in relatively few words. CoMem uses a *storytelling metaphor* for the evolution history explorer. The CoMem Evolution History Explorer (EHE) module enables the designer to explore the evolution history of any item

selected from the overview - from an abstract idea to a fully designed and detailed physical component, discipline subsystem, or even entire project. (Figure 1c)

We find that CoMem as a tool allows users to visualize large hierarchies at a glance as well as provides the functionality to explore local context for any part of that large hierarchy. Hence CoMem addresses the content and context based limitations in the process of knowledge reuse.

3. CoMem Exploration Framework

CoMem provides a systematic framework for exploration in the process of knowledge re-use that compliments the simple search and retrieval paradigm provided by traditional tools that results in a more detailed and well-informed final design solution. For the purpose of our discussion we will consider that for any two corporate objects A and B, object B is relevant to object A if: (1) the user is currently working on object A and object B is potentially reusable; and (2) the user is considering reusing object A and object B is somehow related to object A, such that knowledge about object B helps the user understand object A or improve the design of object A.

We can show the importance of such an exploration oriented knowledge-reuse strategy based on the following probabilistic formulation:

We define the function $Q(A|B)$ as follows:

- The designer is working on the design of item A
- B is an item of information or knowledge from the corporate memory that would be of value during the design of A .
- $Q(A|B)$ is some quantification of the probability of finding the item of information B from the corporate memory or a large archive during the design of A .

For example:

$Q(\text{ClarkCenterSteelFrame} | \text{FireProtectionManual})$

is the probability of being able to retrieve a manual on fire protection while a designer is working on a steel frame for the Clark Center project. For a novice designer this probability would be quite small, because a novice might not know that fire protection is a major issue in the design of steel structures, and the terms fire protection might not appear anywhere in the technical documents for the steel frame.

Similarly, $Q(\text{ClarkCenterSteelFrame} | \text{SteelDesignManual})$ is the probability of being able to retrieve a manual on steel design while a working on the steel frame for the Clark Center project.

$Q(\text{SteelDesignManual} | \text{FireProtectionManual})$ is the probability of being able to retrieve a manual on fire protection while reviewing the manual on steel design.

Intuitively, the following will be true:

$$Q(\text{ClarkCenterSteelFrame} | \text{FireProtectionManual}) \ll Q(\text{ClarkCenterSteelFrame} | \text{SteelDesignManual}) \times Q(\text{SteelDesignManual} | \text{FireProtectionManual})$$

In other words, a novice designer working on a steel frame is much more likely to find a manual for steel design and *then* find a manual on fire protection than he or she would be to find a manual on fire protection *directly* while working on the steel frame.

The above example emphasises the importance of *contextual exploration* in the knowledge reuse process. In a large repository such as a corporate memory, a valuable item might be beyond the reach of a query such as that supported by traditional retrieval tools. However, by supporting *exploration*, rather than *retrieval*, valuable items can be found in large archives despite the flaws of modern information retrieval systems and the inexperience of the designer. This is especially true for *large* repositories. Figure 2 conveys this premise. For small repositories, traditional tools may be just as effective as or even slightly more effective than CoMem, but their efficacy rapidly declines as the repository size increases.

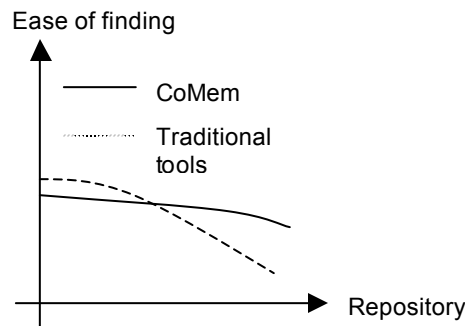


Figure 2: Performance of traditional tools decreases considerably with an increase in repository size, while the exploration framework provided by CoMem ensures an effective knowledge reuse process.

3.1 CoMem-ThinkTank Testbed

To demonstrate the value of CoMem for exploration in the knowledge reuse process requires a testbed made of a database of appreciable size comprising of previous projects of a corporation. The ThinkTankTM archive is one such database, that evolved in real-time over a period of 4 years. Specifically, ThinkTank is a web-based collaboration tool, developed at the PBL Lab at Stanford that provides many of the advantages of face-to-face team meetings combined with the utility of e-mail and the organization of database technology (Fruchter et al 2003). ThinkTank facilitates AEC project team members to create, capture, share, email, reply, attach additional documents, track, sort, search, archival, and re-use data, information, and knowledge. ThinkTank functionalities evolved based on ethnographic studies of the state-of-practice in industry and academia focused on the cross-disciplinary project team communication needs.

ThinkTank uses Microsoft AccessTM as the underlying database that is organized by project groups, with each project group having a private, password-protected, virtual project space. Within the project space the content is organized based on the “book” metaphor with three levels of organization: *Forums* representing key project disciplines or project phases for instance similar to book chapters, *Subjects* representing or discussing component design issues brought up within a *forum* or a *discipline*, similar to section within a chapter of a book, and threaded *Messages* representing the actual ideas captured and shared among team members, similar to the paragraphs within a section of a book (Figure 3). Messages are created as needed over the course of the project by the team members. Each group can create any number of forums, within each forum there can

be any number of subjects, within subjects any number of threads, and within threads any number of messages. Within the context of ordinary project communications, there are no practical limits to the numbers of each.

To provide rapid navigation, search, exploration, and knowledge re-use we have integrated ThinkTank with CoMem (Figure 4)

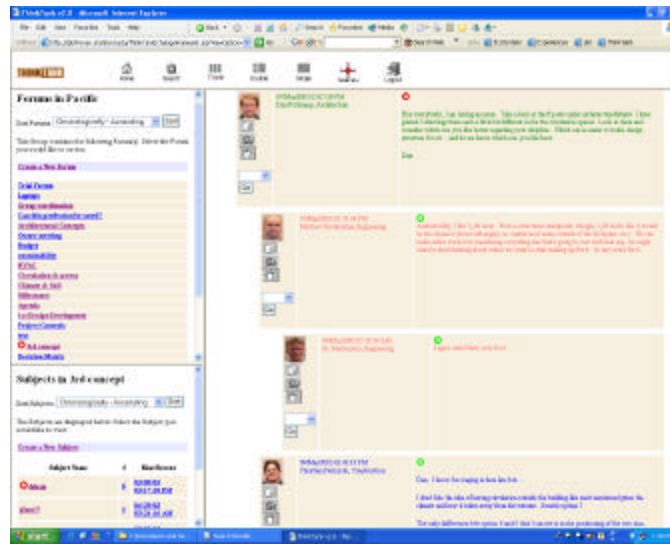


Figure 3: Screen snapshot ThinkTank™ of the Pacific2003 team discussing issues about the first floor plan of their building. Architect Dan from Kansas University, Structural Engineers Matt from Stanford and Bo from KTH Sweden Construction Managers Christian from Stanford University and Andrea from Bauhaus University Germany

The resulting CoMem-ThinkTank testbed, consists of an archive that is constantly growing and currently includes 64 projects over a period of 4 years. These 64 projects comprise of 780 disciplines and approximately 2400 components with 8000 versions in all.

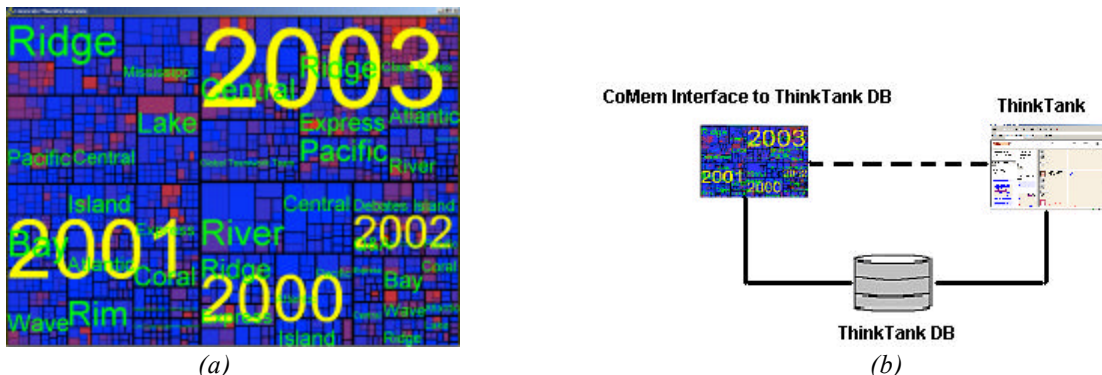


Figure 4: The CoMem as the front-end to ThinkTank database: (a) CoMem Overview Map, (B) CoMem-ThinkTank

3.2 Formal Exploration Framework

We formalize the notion of exploration and classify exploration activity in the following two categories:

- *Breadth-Only Overview Exploration* that facilitates localization of pockets of re-usable knowledge from the Corporate Memory
- *Iterative Breadth-Depth Exploration* that helps identify those re-usable components of the Corporate Memory that may not necessarily be contextually relevant to the problem at hand.

The exploration activity in CoMem is initiated through a user text query. Based on the query a color-coded map is computed through the LSI engine and presented to the user (Figure 5). On this map, all items are coded in colors ranging between red and blue in decreasing order of contextual relevance to the original query, red items being the most relevant and blue items being the least relevant. This relevance score is obtained through the latent semantic indexing algorithm applied to the document corpus of the archive.

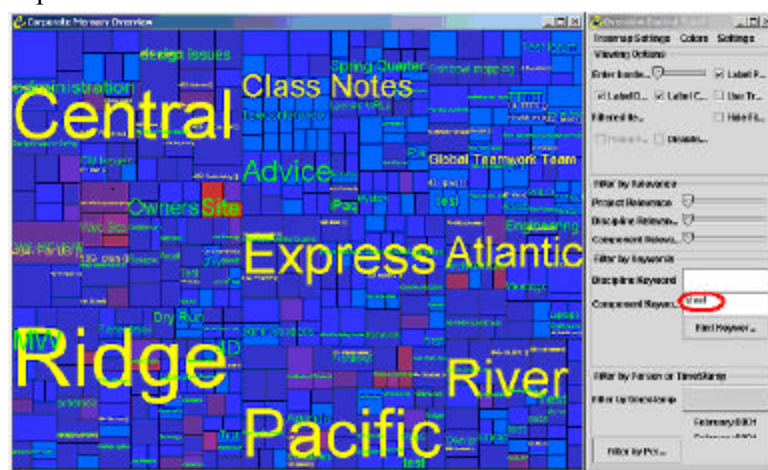


Figure 5: User types in a query, e.g., “steel” and is presented with a CoMem Overview Map highlighting items of high relevance in the corporate memory

The first type of the exploration activity that a designer can carry out is the *Breadth-Only Overview Exploration*. Such exploration activity is appropriate when the designer is seeking extensive information and insight on multiple knowledge areas. The designer initiates this activity by submitting to CoMem, queries for each of the knowledge areas that he is interested in exploring and correspondingly obtains differently color coded overview maps for each of these queries. Each map identifies parts of the Corporate Memory that would be of interest to the designer in exploring the aforementioned knowledge areas. Based on these maps the designer can identify projects and disciplines of the Corporate Memory that reference all or most of the different knowledge areas he is interested in exploring and hence carry out localized exploration of those projects or disciplines. Having found such projects and disciplines of the Corporate Memory that reference all the queried knowledge areas, the designer can further explore these through the CoMem prototype, corporations project archives or feedback from an expert in the corporation who has worked on these projects and disciplines.

Assume a designer seeks insight in knowledge areas X, Y and Z. He creates queries for each of these areas and gets 3 maps correspondingly as shown in Figure 6. The designer finds projects A, C and D relevant to knowledge area X, projects A, B and C relevant to knowledge area Y, and projects B, C and E relevant to knowledge area Z. The designer localizes his exploration to project C as it references all the three knowledge areas he is interested in and goes through all the information available through the corporate memory. The designer further speaks to his senior at the firm who has worked on project C and obtains the necessary insight he seeks (Figure 6)

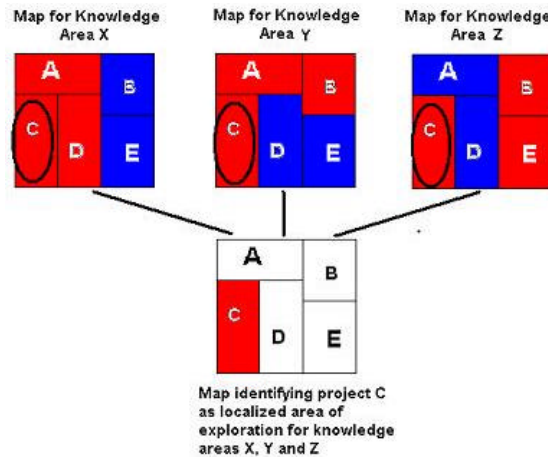


Figure 6: Maps corresponding to queries for knowledge areas X, Y and Z. Map showing localized area of search as project C for all knowledge areas of interest X, Y and Z.

The other type of exploration activity that a designer can carry out is an *iterative breadth-depth exploration*. Such exploration is appropriate in a situation where a designer has a target end component design but is not aware of all design aspects that need to be considered for this component. The designer first formulates his query based on all knowledge areas associated with the design that he is aware of. Based on the query CoMem will present to the designer a color-coded map that identifies all items in the corporate memory that are relevant to the knowledge areas represented through the query. The designer then explores the context of the most relevant items in the map through the project context explorer (PCE). As mentioned earlier, the item of interest selected from the map becomes the focal node in the project context explorer, with all relevant items organized in decreasing order of structural distance (defined as 1 if item is in same project and discipline as focal item, 2 if the item is in the same project but a different discipline and 3 if both items are in different project and disciplines) and contextual relevance. This organization of the project context explorer provides the designer with the opportunity to carry out a detailed exploration of the context of each relevant item. The designer can browse through the different versions associated with any item of interest within the EHE that displays all the versions on an item being invoked from the PCE. This constitutes the depth-based exploration in this process. Based on information acquired in this manner the designer can decide to further explore the context of an item that he finds interesting. Invocation of any item within the PCE presents to the designer a new PCE with the invoked item as the focal node and its corresponding context. This constitutes the breadth-based exploration in this process.

Such interwoven breadth-depth exploration allows the designer to unearth knowledge areas and design aspects associated with the target design that was unaware of originally. The user is made aware of such unknown design aspects as a result of contextual exploration of items that represent entry points to contexts tangential to the context being explored by the user. Typically such items have medium to high contextual relevance and a high structural distance from the focal node. Substantive exploration of such tangential contexts through the project context explorer informs the user of different knowledge areas and design aspects that he had not considered in his original query.

For instance, consider a designer who formulates a query to design a component X. CoMem returns a map that identifies item A with highest relevance to the query. Designer invokes the PCE with

component A as the focal node and explores P - the context of item A. The designer finds item E interesting as it represents a design aspect he had not considered initially and explores its design versions from the EHE. The designer further explores the context Q of item E by invoking a new PCE with item E as the focal node. In this manner the designer iteratively continues to explore tangential contexts Q and R of items E and J respectively, each unearthing new design aspects and knowledge areas associated with the designer's task. Finally the designer concludes his exploration with depth exploration of the various versions of item N and re-thinks his design and query (Figure 7).

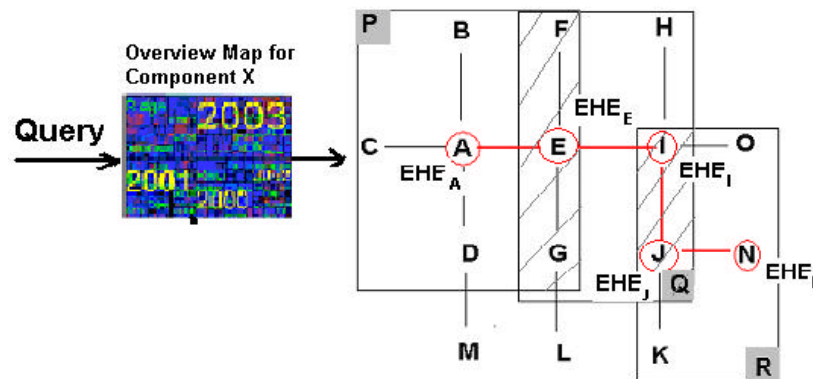


Figure 7: Iterative breadth-depth exploration of adjoining local contexts with shared components revealing new design aspects that the designer was unaware of at the start of the exploration. The hashed areas represent the shared items in the contexts P and Q of items A and E, and the shared items in the contexts Q and R of items E and J, respectively.

4. Exploration and Discovery Scenario

Joe is a novice engineer in a mid size structural design firm. He is working on his first building project - a Hotel in Sacramento. He was assigned to work on a steel structure alternative that would address the architect's concept for the hotel and the overall "sustainable design" objective of the project. The architect's concept is to have a large atrium in the center of the building and use exposed a MEP system to give the building a mechanical toy look and feel. Joe decides to explore some of the firm's past projects and find reusable concepts and details for steel structure schemas, atrium solutions and exposed MEP approaches. Over the years the firm worked on more than fifty projects and has archived in ThinkTank the concepts, solutions, interactions and communications that led to decisions.

Joe decides to start with a breadth-only search and identify the key projects he could further explore in detail. For this purpose Joe opens ThinkTank and starts CoMem engine to perform separate searches for the following topics: *LEED* for sustainability information since this is his first time he tackles this topic (Figure 8a), *steel structure* (Figure 8b), *atrium* (Figure 8c), *Exposed MEP* (Figure 8d). For each search CoMem returns a color-coded map of the large corporate archive indicating the potential projects (areas on the map) that are relevant to his specific search. The color-coding scheme uses red for the highest relevance related to the query, blue for the lowest or no relevance to the query, purple spectrum for the contextually relevant and related items. Each map has a number of rectangles color-coded in red, which Joe plans to explore. He was able to quickly run the queries and obtain multiple, contextually color-coded Overview Maps that provide him with complete and succinct overviews of the large corporate archive, he can decide at a glance which projects address multiple issues he is interested in and explore those first. In this case Pacific project – has items highlighted in red in two Overview Maps that represent the steel structure and exposed MEP queries, and Ridge project – has items highlighted in red in three Overview Maps that represent the

steel structure, LEED, and exposed MEP queries. Wave project has items highlighted in red in two Overview Maps steel structure and exposed MEP queries, and Atlantic project has items highlighted in red in two Overview Maps steel structure and atrium queries.

Joe decides to explore in detail the projects in the following order Ridge, Express, Pacific, and Wave projects based on the identified overlaps in query results shown in the four Overview Maps. He starts a detailed breadth-depth iterative exploration by using the PCE and EHE of CoMem.

At some point during his exploration, Joe selects steel structure item in the Pacific project in the Overview Map. As a result, the PCE and EHE are updated and the relevant data is retrieved from the database. (Figure 9) As he explores the PCE that has “steel structure” in the focal point he discovers that another relevant item that is colored in purple is labeled “exposed ceiling” and is part of the “fire protection” subject. This is a sideways exploration on a subject that is contextually related to “steel structure.” Since Joe is a novice engineer, he did not know that for steel structures he has to consider fire protection. He decides to learn more about it and explores the threaded discussion items related to “fire protection” and “exposed ceiling” in the EHE. He decides to reshuffle the configuration of the PCE with “exposed ceiling” and with “fire protection” in the focal point. This helps him visualize, explore and discover other topics related to steel structure and exposed ceiling that he was not aware. During this iterative breadth-depth exploration he discovers some interesting structural connection details and valuable pointers to MEP subcontractors that he might contact in case the project team decides to opt for the steel structure solution.

New items become closer to the focal point as their relevance increases with respect to the new point of interest. Consequently, even though Joe did not think of fire protection issues related to steel structures, and the relation between exposed MEP and ceiling, CoMem affords the search and discovery of these key design aspects.

5. Discussion

How do you find the metaphorical needle in a haystack? How do you find a needle in a haystack if you don't even know you're looking for a needle, or if you've never seen a haystack before? How can a computer system support designers of constructed facilities in reusing building designs and design experiences from previous projects? Those are the kind of questions the CoMem (Corporate Memory) technology presented in this paper tries to address. Reusing design knowledge from a corporate archive of past projects is an important process that often fails. We attribute this failure to the fact that state-of-practice archiving systems, e.g., archives of paper drawings or electronic files arranged in folders, do not support the designer in *finding* reusable items and *understanding* these items in context in order to be able to reuse them. We argue that informed *finding*, *understanding*, and *reusing knowledge* go beyond simple search and retrieval and require a capability to support exploration that assists a user to (1) gain a global overview of potentially reusable resources from a large corporate memory, i.e., rapid *breadth-only overview exploration* that facilitates localization of pockets of re-usable knowledge from the Corporate Memory; and (2) identify related but critical knowledge, i.e., *iterative breadth-depth exploration* that helps identify those re-usable components of the Corporate Memory that may not necessarily be contextually relevant to the original query at hand. This in turn will lead to higher quality designs and reduced rework.



(a) Query: LEED



(b) Query: Steel Structure



(c) Query: Atrium



(d) Query: Exposed MEP

Figure 8: Identified Areas of Interest for Exploration in CoMem Color Coded Maps for Different Search Queries.

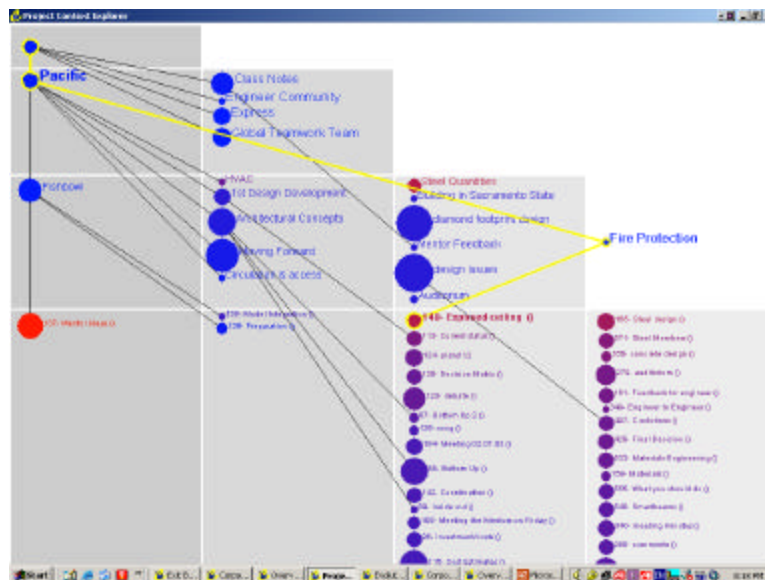


Figure 9: CoMem Project Context Explorer (PCE) with “Steel Structure” item of Pacific Team in the Focal Point. Exploration of Relevant items (purple colored) in the PCE Joe Discovers “Exposed Ceiling” and “Fire Protection” as items related to steel structures that he did not consider before he started his exploration for reusable knowledge.

6. References

- Fruchter, R. and Demian, P. (2002) “*CoMem: Designing an Interaction Experience for Reuse of Rich Contextual Information from a Corporate Memory*,” *AIEDAM* International Special Issue on “Human Computer Interaction in Engineering Context”, guest co-editors Ian Parmee and Ian Smith, 16,127-147
- Fruchter, R. Demian, P, Yin, Z., and Luth, G. (2003), “*Turning AEC Knowledge into Working Knowledge*,” Proc. of Computing in Civil Engineering Conference in conjunctions with ASCE National Convention, ed. Ian Flood, Nashville, 2003.
- Furnas, G. W. (1981). *The FISHEYE View: A New Look at Structured Files*. In Readings in Information Visualization, (Card, S. K., et. al., Eds.), pp. 312-330. Morgan Kaufmann, Los Altos, California.
- Lamping, J. and Rao, R. (1995) *The Hyperbolic Browser: a Focus + Context Technique for Visualizing Large Hierarchies*, in Readings in Information Visualization, (Card, S. K., et. al., Eds.), pp. 383-408. Morgan Kaufmann, Los Altos, California.
- Landauer, T. K., & Dumais, S. T. (1995). *A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge*. *Psychological Review*, 104, 211-240
- Shneiderman, B. (1994). *Dynamic Queries for Visual Information Seeking*. In Readings in Information Visualization, (Card, S. K., et. al., Eds.), pp. 236-243. Morgan Kaufmann, Los Altos, California.
- Shneiderman, B. (1999). *Supporting Creativity with Advanced Information-Abundant User Interfaces*. In Human-Centred Computing, Online Communities, and Virtual Environments, (Earnshaw, R., et. al., Eds.), pp. 469-480. Springer-Verlag, London.